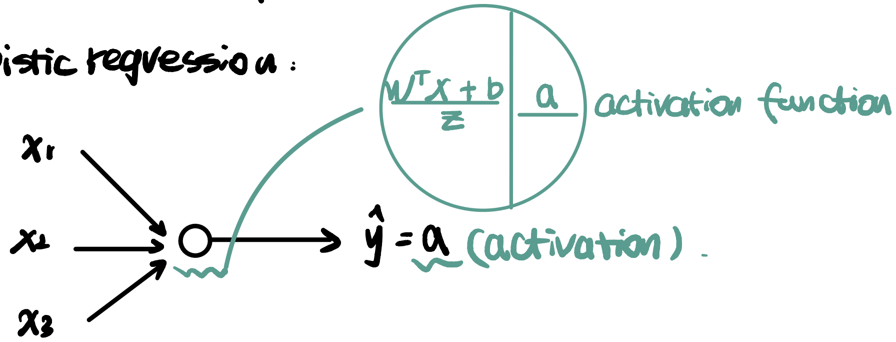
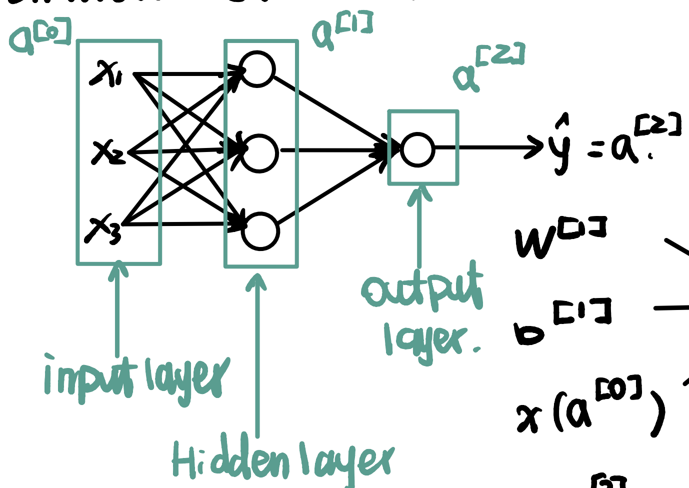
Shallow Neural Network

An example view of a shallow neural network.

logistic regression:



Shallow Neural Network:



$W^{[1]}$

$b^{[1]}$

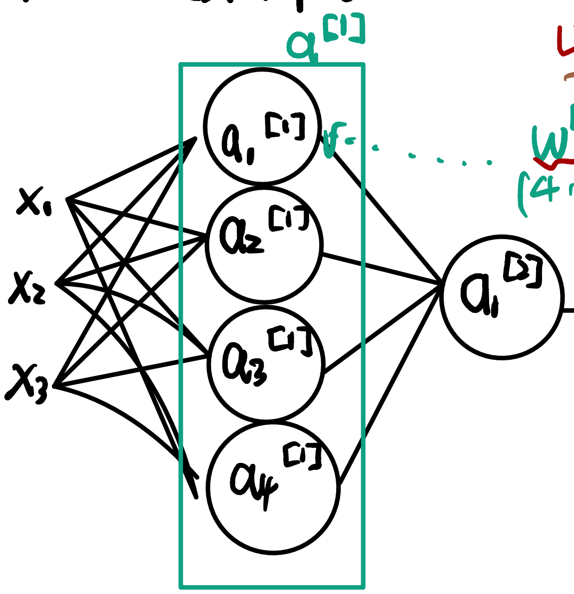
$x\,(a^{[0]})$

$\rightarrow z^{[1]} = W^{[1]}\boxed{x} + b^{[1]} \rightarrow a^{[1]} = \sigma(z^{[1]})$

also can be noted as $a^{[0]}$ (input).

$\longrightarrow z^{[2]} = W^{[2]}a^{[1]} + b^{[2]} \rightarrow a^{[2]} = \sigma(z^{[2]}) \rightarrow L(z^{[3]}, y)$

Another example:



$\rightarrow \hat{y} = a_1^{[2]}$

4 sets of weights w_i, where $w_i \in \mathbb{R}^{③} = n_x$

Think about them as 3, 4-dimension vectors, where V_1 contains weight for X_1 from all 4 hidden neurons, etc.

$W^{[1]}$, $b^{[1]}$

$(4, 3)$ $(4, 1)$

4 sets of bs

Also one set of bs: b_1, b_2, b_3, b_4, one in each hidden neuron

$$a^{[1]} = \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \\ a_3^{[1]} \\ \vdots \\ a_4^{[1]} \end{bmatrix}$$

The above neural network is usually refered as "2 layer NN" (input layer's not counted).

Computing Neural Network output:

w_i, x are all col vectors.
w_i^T becomes a row vector.
$w_i^T \cdot x$ is a scalar.

$$z_1^{[1]} = w_1^{[1]T}x + b_1^{[1]}, \quad a_1^{[1]} = \sigma(z_1^{[1]})$$
$$z_2^{[1]} = w_2^{[1]T}x + b_2^{[1]}, \quad a_2^{[1]} = \sigma(z_2^{[1]})$$
$$z_3^{[1]} = w_3^{[1]T}x + b_3^{[1]}, \quad a_3^{[1]} = \sigma(z_3^{[1]})$$
$$z_4^{[1]} = w_4^{[1]T}x + b_4^{[1]}, \quad a_4^{[1]} = \sigma(z_4^{[1]})$$

$$z^{[1]} = \begin{bmatrix} — w_1^{[1]T} — \\ — w_2^{[1]T} — \\ — w_3^{[1]T} — \\ — w_4^{[1]T} — \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \\ b_4^{[1]} \end{bmatrix} = \begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \\ z_3^{[1]} \\ z_4^{[1]} \end{bmatrix}$$

As noted in prev page, weight matrix W is implemented as the following:

- i^{th} row vector: a set of weights in the i^{th} neuron in this hidden layer.
- j^{th} col vector: a set of weights in each neuron for $\underline{x_j}$

$$a^{[1]} = \begin{bmatrix} a_1^{[1]} \\ \vdots \\ a_4^{[1]} \end{bmatrix} = \sigma(z^{[1]})$$

vectorization:

Given Input x:

$\rightarrow \underset{(4,1)}{z^{[1]}} = \underset{(4,3)}{W^{[1]}} \underset{(3,1)}{\boxed{a^{[0]}}} + \underset{(4,1)}{b^{[1]}}$

$\rightarrow \underset{(4,1)}{a^{[1]}} = \sigma(\underset{(4,1)}{z^{[1]}})$

$\rightarrow \underset{(1,1)}{z^{[2]}} = \underset{(1,4)}{W^{[2]}} \underset{(4,1)}{a^{[1]}} + \underset{(1,1)}{b^{[2]}}$

$\rightarrow \underset{(1,1)}{a^{[2]}} = \sigma(\underset{(1,1)}{z^{[2]}})$

$a^{[0]}$.shape = $\underline{(n_x, 1)}$

this is just 1 training sample.

Next step: vectorizing across multiple examples.